

## 科学研究費助成事業 研究成果報告書

令和 2 年 7 月 8 日現在

機関番号：32692

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K01134

研究課題名(和文)ディープラーニングを用いた教育ビッグデータ解析による要ケア学生の早期抽出の実践

研究課題名(英文)Practice of Early Detection of Care Needed Students by Analyzing Educational Big Data Using Deep Learning

研究代表者

安藤 公彦(ANDO, Kimihiko)

東京工科大学・先端教育支援センター・講師

研究者番号：00551863

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：要ケア学生抽出のための教育ビッグデータ収集および統合システムの開発を行った。座席および時刻付き出席データは年間100万件、学習管理システムの動作ログは年間2000万件収集している。

深層学習を用いることで会話を高い水準で多次元に分類することが可能とする手法を開発した。しかし、この手法による分類は教師データとして利用した元データの影響が大きく、講義が異なるなどチャットの利用シーンが異なると精度が低くなる性質があり利用シーンごとに教師データを用意する必要があった。そこで、転移学習を利用することで、従来方式より1/10程度の教師データで高い精度で分類できる手法を開発した。

研究成果の学術的意義や社会的意義

大学における要ケア学生抽出のための基盤となる情報収集システムを構築し、従来よりも正確な分析が可能となる詳細な出席データおよびMoodleログの収集を全学的に実際に行なった。これにより、実データによる要ケア学生の分析が可能となり、要ケア学生抽出の実践に向けた準備ができたといえる。

さらに、より詳細な分析を行うための会話データ分類を深層学習により開発をした。これにより複数の次元による分類と、転移学習による少ない教師データによる高精度な分類を可能とした。

研究成果の概要(英文)：We have developed an educational big data collection and integrated system for selecting students who need care. 1 million attendance data with time and seats number and 20 million Learning Management System's operation logs are collected annually.

We have developed a method that enables high-level multi-dimensional classification of conversations by using deep learning. However, the classification by this method has a large influence of the original data used as teacher data, and the system becomes low when the usage scene of chat such as different lectures is different, so it was necessary to prepare teacher data for each usage scene.

.. Therefore, by using transfer learning, we have developed a method that can classify with high accuracy with about 1/10 of the teaching data compared with the conventional method.

研究分野：教育工学

キーワード：Learning Analytics Institutional Research 深層学習 学習管理システム

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

近年教育分野においても学習管理システム Learning Management System(以下 LMS)の導入や各種情報の IT 化により、ビッグデータの活用が進んできている。例えば Learning Analytics(LA)の分野においては、LMS の課題の内容・進行状況や各種ログから、各学生の学習状況を把握することで、講義進行や講義内容の動的な変更や、グループ学習におけるグループ状態の可視化などが可能となっている[1]。また Institutional Research(IR)の分野においては、各種データを統合的に分析することで、大学全体の状況の可視化や、各学部の差異など多くの分析が可能となってきた。

しかし、IR においては各種情報の統合が課題となっており、どのような仕組みで大学中の情報を取得し分析できる環境を構築するかが重要であると考えられる。

また、一方で大学における重要な課題の一つとして、何らかの問題を抱えているため、授業への不参加、ひいては退学へと追い込まれてしまう学生の発見とそのケアが挙げられる。

### 2. 研究の目的

従来、ケアが必要な学生の発見は、各講義の出席状況や成績の取得状況等から、ある程度把握されてきていた。しかし、成績が悪くなった時点ですでにケアするには遅すぎる場合が多く、それは出席状況が悪くなってからでも同じである。理想としては、出席状況が悪くなる前にケアすべき学生の抽出をすべきである。特に、入学時には問題がなく、学生生活を続けていくうちにケアが必要となる学生の検出が必要とされる。そのためには、単なる出席や成績より細かな情報が必要と考えられる。

本学では、教育系システムが自作のため、本学独自のより詳細な出席システムがあり、どの座席からいつ出席登録したかを記録できる。そのためより詳細な出席状況の把握が可能である。また、LMS として Moodle を導入しており、毎学期、全講義のコースを作成し、すべての講義での Moodle 利用を推奨している。そのため、学生の学習状況についてもより詳細に分析することが可能であると考えられる。

しかし、定量的なデータだけでは十分な分析は難しく定性的なデータの分析が必要となる。本研究では提出された自由記述の課題の回答やチャットデータから、学生の状況を把握できるのではないかと考え、その分析手法の開発を目指す。

### 3. 研究の方法

#### (1) 組織内 IT サービス

大学に限らず組織内の IT サービスは、外部ベンダーが提供するサービスを利用する方式(外注)と、組織内でサービスを自前で構築する方式(自前)が考えられる。一般的には外注が安定性の観点から推奨される。自作する場合は予算の都合によるものがほとんどである。しかし、外部仕様がまだ確定していなかったり、仕様が頻繁に変更されたりする場合には、自作でサービスを構築することのメリットが生まれる。

大学での IT サービスを考えると、要求仕様がはっきりしていてかつ安定性と安全性が必要とされる業務系のサービスは外注が望ましい。例えば教務システムや図書館システムである。一方、教員や学生が直接操作する LMS などの教育系サービスは、授業形態や利用のされ方により、常に要求が変わるため、自作が望ましいといえる。

そこで本学では IT サービスを業務系と教育系に分け、業務系は外注のサービス、特にメンテナンス性と安定性の観点からクラウドサービスを導入した。一方、教育系はオープンソース LMS である Moodle に自作のモジュールを追加して利用し、出席登録システムを兼ねる学生ポータルを自作した[2]。図 1 にその概略を示す。教育系に関しては応答性や通信料を踏まえ、オンプレミスのサーバーで運用をしている。

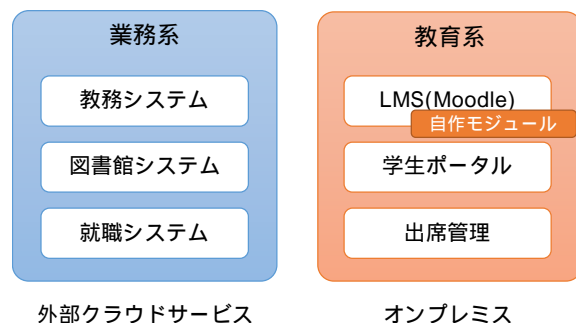


図 1 学内 IT サービス環境

#### (2) 教育系システムの開発と導入

要ケア学生の抽出に必要な情報として出席データと学習データが挙げられる。出席データは要ケア学生の抽出に最もよく使われるデータであるが、単純な出欠席の情報だけでは学生の状況を知るには不十分であり、欠席が多くなった状態にならなければ抽出は不可能である。しかし、

欠席が増えた状況はすでに早期ケアを行う時期を過ぎているといえる。そこで本研究では座席に番号を振りコンピュータおよびスマートフォンから学生が登録することで出席を可能とする「Web 出席システム」を開発し全学に導入し、出席時間および出席位置まで記録することとした。次に学習データであるが、講義中であれば教員の目視によりある程度学習状況の把握は可能である。しかし、学外での学習状況の把握は課題の内容や提出状況程度しか把握することはできず、その全貌を明らかにすることは難しかった。しかし、近年、LMS の導入により、より詳細な学習状況の把握が可能となってきた。本研究ではなるべく多くの学習データを取得するため、学習管理システム Moodle を導入し、全講義に対し専用のコースを提供する。また、膨大な量となる Moodle のデータは、学修ポートフォリオへの転用も考え、全て利用可能な状態で保存をする。

### (3) 情報統合手法

外部の各種サービスを導入している組織において、すべてのサービスのデータを統合することは、すべての外部サービスを同一ベンダーにする以外は非常に困難である。本学では、業務系の各種サービスをクラウド化するとともに、異なるベンダー間のサービスにおいてもデータの統合を可能とするための中核 DB を設置した。

図 2 に中核データベース(以下 DB)の動作を示す。中核はクラウド上にある DB サーバーであり、各ベンダーに業務系各種サービスの DB のレプリカをこの中核 DB 上に設置するよう依頼することで、各種サービスのデータを収取できるようにしたものである。どのテーブルのどの列がどのような意味であるか適切に管理することで、サービス間のデータの統合が可能となる。基本的に学生に関するデータは学籍番号をキーとして統一的に扱うことが可能である。

一方、教育系サービスは自作のため、中核 DB へのデータのアップロードは容易であり、中核 DB 内のデータ構造に合わせてシステムを構築することが可能となっている。

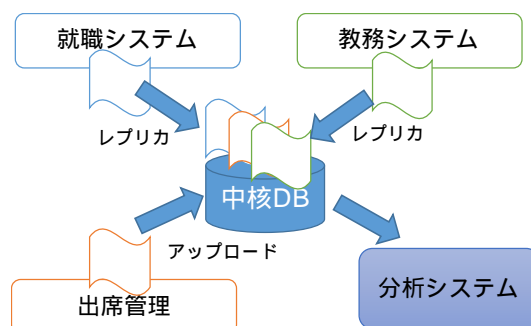


図 2 中核 DB

## 4. 研究成果

### (1) 取得データの推移

図 3 に座席位置および出席時間が記録された出席データ件数の推移を示す。2017 年度以降は 100 万件前後で推移しており、収集すべき出席データは収集できていると考えられる。



図 3 座席および時刻付き出席データ数の推移

図 4 に Moodle の Feedback 回答数を示す。Feedback 回答数は Feedback 単位で数えており、

Feedback の中に設問が複数あったとしても 1 件と数えている。2018 年度を最大に 30 万件程度の回答数がある。

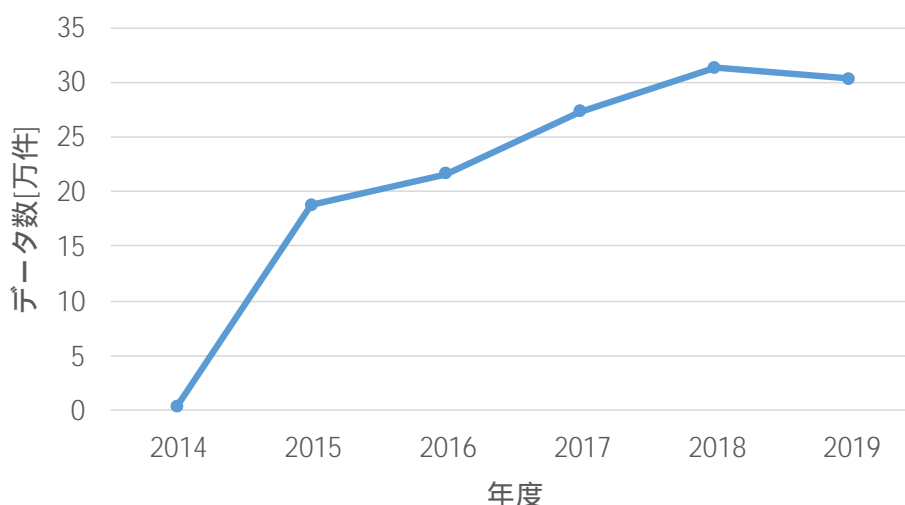


図 4 Moodle の Feedback 回答数

図 5 に Moodle のログデータ件数を示す。ログデータ数は、ユーザの操作ログである origin が web のものだけを抽出している。右肩上がりに増えており、これは Moodle の利用頻度が増加傾向になることを示している。2019 年度で 2 千万件を超えている。

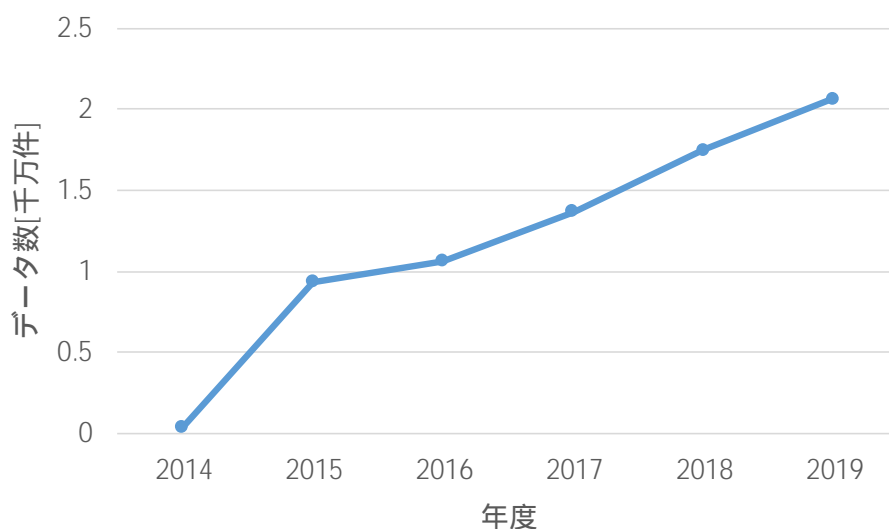


図 5 Moodle のログ件数

## (2) 文章データ分類の深層学習の手法の開発

文章の分析には深層学習を用いる。チャットデータの分析として Seq2Seq を用いた手法を開発した。この手法を用いることで、会を高水準で多次元に分類することが可能となった。しかし、この手法による分類は教師データとして利用した元データの影響が大きく、講義が異なるなどチャットの利用シーンが異なると制度が低くなる性質があり利用シーンごとに教師データを用意する必要があった。そこで、転移学習を利用することで、従来方式より 1/10 程度の教師データで高い精度で分類できる手法を開発した[3]。

本手法では文章を分類するが異なる性質を持つ 4 つの次元についてそれぞれ分類した。また次元によって分類数は 3-5 と異なる。係数は分類結果の精度を表すが、これが高いほど精度が高い。一般的に深層学習においては教師データ数が多いほど精度が高くなる。ある係数を達成するのに必要な教師データ数を従来手法と本手法で比較をした。その結果が表 1 である。従来手法で必要とされる教師データ数を 100%とした場合の本手法で必要とされる教師データ数である。なお、教師データ数が足りず従来手法では到達できなかった精度は「-」で示す。最大で 9%まで縮小できている。

表1 ある精度を達成するために必要とする教師データ数（従来方式との比率）

係数	次元1	次元2	次元3	次元4
0.75	44%	-	85%	-
0.7	9%	59%	79%	-
0.65	8%	28%	93%	37%
0.6	11%	11%	81%	37%

(3) 今後の展望

要ケア学生の抽出には、出席情報や LMS のログに加え、授業評価アンケート結果や各種業務系サービスの情報など、多岐にわたるデータ収集及び分析が必要となる。データの収集は中核 DB を中心とし、各種サービスのデータの統合を行い、すでにビッグデータといってよい量が収集されている。出席情報からは教壇からの位置関係や隣接する学生の情報などから、今までにない分析が可能になると期待できる。また、文章データの分析手法を用いてより定性的な分析が可能になると考えらる。

参考文献

[1] 安藤公彦、柴田千尋、稲葉竹俊，深層学習技術を用いた自動コーディングによる協調学習のプロセスの分析，コンピュータ&エデュケーション,43,pp79-84

[2] 安藤 公彦，アクティブラーニング全学導入にむけた I C T 基盤の構築と実践,平成 29 年度教育改革 ICT 戦略大会

[3] 守屋 俊，柴田 千尋，安藤 公彦，稲葉 竹俊，協調学習における会話分析用教師データの削減を可能とする転移学習の活用，コンピュータ&エデュケーション,47,pp43-48

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Kimihiro Ando, Chihiro Shibata and Taketoshi Inaba	4. 巻 3,4
2. 論文標題 Coding Collaboration Process Automatically: Coding Methods Using Deep Learning Technology	5. 発行年 2017年
3. 雑誌名 International Journal of Advances in Intelligent Informatics	6. 最初と最後の頁 345-354
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 守屋 俊, 柴田 千尋, 安藤 公彦, 稲葉 竹俊	4. 巻 47
2. 論文標題 協調学習における会話分析用教師データの削減を可能とする転移学習の活用	5. 発行年 2019年
3. 雑誌名 コンピュータ&エデュケーション	6. 最初と最後の頁 43-48
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Takahiro Kanayama, Kimihiro Ando, Chihiro Shibata, Taketoshi Inaba
2. 発表標題 Using Deep Learning Methods to Automate Collaborative Learning Process Coding Based on Multi-Dimensional Coding Scheme
3. 学会等名 eLmL 2018（国際学会）
4. 発表年 2018年

1. 発表者名 Chihiro Shibata, Kimihiro Ando, Taketoshi Inaba
2. 発表標題 Towards Automatic Coding of Collaborative Learning Data with Deep Learning Technology
3. 学会等名 eImL2017（国際学会）
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	柴田 千尋  (SHIBATA Chihiro)  (00633299)	東京工科大学・コンピュータサイエンス学部・講師   (32692)	
研究分担者	田胡 和哉  (TAGO Kazuya)  (10188229)	東京工科大学・コンピュータサイエンス学部・教授   (32692)	
研究分担者	稲葉 竹俊  (INABA Taketoshi)  (10386766)	東京工科大学・教養学環・教授   (32692)	